

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:

Takuya NODA, et al.

Application No.:

Group Art Unit:

Filed: October 20, 2003

Examiner:

For: VOICE INTERACTIVE SYSTEM AND METHOD

**SUBMISSION OF CERTIFIED COPY OF PRIOR FOREIGN
APPLICATION IN ACCORDANCE
WITH THE REQUIREMENTS OF 37 C.F.R. § 1.55**

Commissioner for Patents
PO Box 1450
Alexandria, VA 22313-1450

Sir:

In accordance with the provisions of 37 C.F.R. § 1.55, the applicant(s) submit(s)
herewith a certified copy of the following foreign application:

Japanese Patent Application No(s). 2002-306299

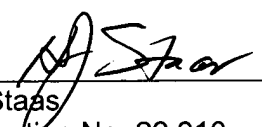
Filed: October 21, 2002

It is respectfully requested that the applicant(s) be given the benefit of the foreign filing
date(s) as evidenced by the certified papers attached hereto, in accordance with the
requirements of 35 U.S.C. § 119.

Respectfully submitted,

STAAS & HALSEY LLP

Date: October 20, 2003

By: 
H. J. Staas
Registration No. 22,010

1201 New York Ave, N.W., Suite 700
Washington, D.C. 20005
Telephone: (202) 434-1500
Facsimile: (202) 434-1501

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日
Date of Application:

2002年10月21日

出 願 番 号
Application Number:

特願2002-306299

[ST.10/C]:

[JP2002-306299]

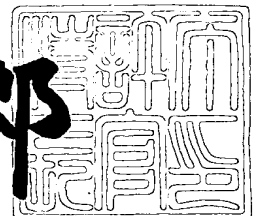
出 願 人
Applicant(s):

富士通株式会社

2003年 3月11日

特 許 庁 長 官
Commissioner,
Japan Patent Office

太田信一郎



出証番号 出証特2003-3015017

【書類名】 特許願

【整理番号】 0295400

【提出日】 平成14年10月21日

【あて先】 特許庁長官 殿

【国際特許分類】 G10L 15/20

【発明の名称】 音声対話システム及び方法

【請求項の数】 7

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

【氏名】 野田 拓也

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

【氏名】 鷲尾 信之

【特許出願人】

【識別番号】 000005223

【氏名又は名称】 富士通株式会社

【代理人】

【識別番号】 110000040

【氏名又は名称】 特許業務法人池内・佐藤アンドパートナーズ

【代表者】 池内 寛幸

【電話番号】 06-6135-6051

【手数料の表示】

【予納台帳番号】 139757

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0115801

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声対話システム及び方法

【特許請求の範囲】

【請求項 1】 入力された音信号に対して音響的な信号処理を行う音響処理部と、

音響的な信号処理後の前記音信号に含まれる音声の内容を認識する音声認識部と、

前記音声の内容に基づいて、音声出力あるいは音声出力と他の情報伝達手段との併用によって、ユーザへ情報を伝達する音声対話部と、

情報を伝達する途中に、前記音響処理部の入力又は出力あるいは外部入力からの入力信号により、情報の伝達を中止するバージイン機能を有するバージイン制御部を含み、

前記バージイン制御部が、前記音響処理部の入力又は出力あるいは外部入力からの入力信号から 1 つ以上の特徴量を検出し、前記 1 つ以上の特徴量に基づいて前記バージイン機能の有効又は無効を決定することを特徴とする音声対話システム。

【請求項 2】 前記 1 つ以上の特徴量のうち少なくとも 1 つが雑音特徴量であり、前記雑音特徴量が所定のしきい値を越えた場合に前記バージイン機能を無効にする請求項 1 に記載の音声対話システム。

【請求項 3】 前記 1 つ以上の特徴量のうち少なくとも 1 つがユーザ音声の S / N であり、前記 S / N が所定のしきい値を越えた場合に前記バージイン機能を有効にする請求項 1 に記載の音声対話システム。

【請求項 4】 前記 1 つ以上の特徴量のうち少なくとも 1 つが前記外部入力からの入力信号から検出された前記ユーザの位置情報であり、前記位置情報と環境情報に基づいて前記ユーザがいる地点の環境評価値を算出し、前記環境評価値が所定のしきい値を越えた場合に前記バージイン機能を無効にする請求項 1 に記載の音声対話システム。

【請求項 5】 前記音声対話部が、音声あるいは他の情報伝達手段の少なくとも 1 つ以上を用いて、前記バージイン機能の有効または無効の状態をユーザに通

知する請求項 1 から 4 のいずれか一項に記載の音声対話システム。

【請求項 6】 入力された音信号に対して音響的な信号処理を行う第一の工程と、

音響的な信号処理後の前記音信号に含まれる音声の内容を認識する第二の工程と、

前記音声の内容に基づいて、音声出力あるいは音声出力と他の情報伝達手段との併用によって、ユーザへ情報を伝達する第三の工程と、

情報を伝達する途中に、前記第一の工程における入力又は出力あるいは外部入力からの入力信号により、情報の伝達を中止するバージン機能を有する第四の工程を含み、

前記第四の工程において、前記第一の工程における入力又は出力あるいは外部入力からの入力信号から 1 つ以上の特徴量を検出し、前記 1 つ以上の特徴量に基づいて前記バージン機能の有効又は無効を決定することを特徴とする音声対話方法。

【請求項 7】 入力された音信号に対して音響的な信号処理を行う第一のステップと、

音響的な信号処理後の前記音信号に含まれる音声の内容を認識する第二のステップと、

前記音声の内容に基づいて、音声出力あるいは音声出力と他の情報伝達手段との併用によって、ユーザへ情報を伝達する第三のステップと、

情報を伝達する途中に、前記第一のステップにおける入力又は出力あるいは外部入力からの入力信号により、情報の伝達を中止するバージン機能を有する第四のステップを含む音声対話方法を具現化するコンピュータ実行可能なプログラムであって、

前記第四のステップにおいて、前記第一のステップにおける入力又は出力あるいは外部入力からの入力信号から 1 つ以上の特徴量を検出し、前記 1 つ以上の特徴量に基づいて前記バージン機能の有効又は無効を決定することを特徴とするコンピュータ実行可能なプログラム。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は、ユーザによる音声入力の内容に応じて、音声出力あるいは音声出力と他の情報伝達手段との併用によってユーザへ情報を伝達する音声対話システムに関する。特に、音声出力あるいは音声出力と他の情報伝達手段との併用によってユーザへ情報を伝達している途中において、ユーザ音声による割り込みがあった場合、情報伝達を中断してユーザによる割り込み音声入力について処理するバージイン機能を備えた音声対話システムに関する。

【 0 0 0 2 】

【従来の技術】

コンピュータ技術の急速な進展に伴って、音声信号の処理に関する技術も長足の進歩を遂げている。それに伴って、インターネット等を介して急速に普及しているボイスポータルに代表される音声対話システムでは、ユーザとシステムとが擬似的な対話を行うことによって、ユーザが所望の情報を合成音声等の音声出力を介して提供している。また、次世代携帯電話や携帯端末（PDA等）では、音声信号に加えて画像情報も扱うことができるため、音声と画像を組み合わせたマルチメディア情報を提供する音声対話システムについても、今後の進展が期待されている。

【 0 0 0 3 】

近年、市場に普及しつつあるボイスポータルは、ユーザビリティを向上させる目的で、ポータルサイト（音声対話システム）からの音声案内の途中であっても、ユーザが割り込みをかけて発声することが可能なバージイン機能が搭載されている場合が多い。当該バージイン機能は、ユーザによる音声入力通信回線等を経由してシステムに入力されることにより、ユーザによる音声入力の開始を検出し、音声出力による案内等を中断して、ユーザによる音声入力を促すものである。ここで、システムへの入力がユーザ音声のみであれば、例えば入力のパワー変動等を監視することによって正確にユーザによる音声入力の開始を検出することができ、音声出力による案内等を中断することが可能であることから、バージイン機能は正常に動作する。

【 0 0 0 4 】

しかしながら、実際にはシステムからの音声出力による案内等が通信回線系で反射して戻る回線エコーや、ハンズフリー電話機等におけるシステムの案内音声を送話器から受話器へ回り込む音響エコー、さらにはユーザ周辺環境から混入する定常あるいは非定常な環境雑音や、通信回線系の雑音等が、ユーザ音声に重畳してシステムに入力される。かかる問題を解決するため、一般的には音響処理によるエコー抑制処理や雑音抑圧処理が行われている。

【 0 0 0 5 】

例えば（特許文献 1）においては、入力された音声のスペクトル分析を行うことによって、音声の回り込みによるエコーを除去することができる音声対話システムが開示されている。

【 0 0 0 6 】

【特許文献 1】

特開平 9 - 2 5 2 2 6 8 号公報

【 0 0 0 7 】

【発明が解決しようとする課題】

上述したように、エコーについては、様々な方法を用いたエコーキャンセラにより抑圧するのが一般的であるが、通信回線系によっては完全にエコーを抑圧することができず、エコーが残留するケースが多いことも事実である。

【 0 0 0 8 】

また、雑音についてはノイズキャンセラで抑圧するのが一般的であるが、定常性雑音については効果的に抑圧することができるのに対して、非定常性雑音については抑圧することが困難であるという問題点もあった。

【 0 0 0 9 】

さらに、エコーキャンセラあるいはノイズキャンセラの双方について、その抑圧効果を高めるようにパラメタ調整を行うことが多いが、パラメタ調整を行うことによって、同時にユーザによる音声入力自体を歪ませることとなり、結果的に音声認識率が低下してしまうという問題点も残されていた。

【 0 0 1 0 】

そして、上記のような理由で発生する残留エコーレベルや非定常性雑音レベルが高い場合においては、従来の音声対話システムでは、バーズイン機能を用いることで残留エコーや非定常性雑音をユーザによる音声入力であると誤判断してしまうことから、音声出力による案内等を停止させた上、音声認識が残留エコーや非定常性雑音を誤認識してしまう場合があり、音声対話システムの誤動作を起こす大きな要因の一つにもなっている。

【 0 0 1 1 】

本発明は、上記問題点を解決するために、様々なエコーや雑音に起因する不測のバーズインと、それに伴う誤認識によるシステムの誤動作の両者を抑止することで、ユーザにとって利便性の高い音声対話システム及び方法を提供することを目的とする。

【 0 0 1 2 】

【課題を解決するための手段】

上記目的を達成するために本発明にかかる音声対話システムは、入力された音信号に対して音響的な信号処理を行う音響処理部と、音響的な信号処理後の音信号に含まれる音声の内容を認識する音声認識部と、音声の内容に基づいて、音声出力あるいは音声出力と他の情報伝達手段との併用によって、ユーザへ情報を伝達する音声対話部と、情報を伝達する途中に、音響処理部の入力又は出力あるいは外部入力からの入力信号により、情報の伝達を中止するバーズイン機能を有するバーズイン制御部を含み、バーズイン制御部が、音響処理部の入力又は出力あるいは外部入力からの入力信号から1つ以上の特徴量を検出し、1つ以上の特徴量に基づいてバーズイン機能の有効又は無効を決定することを特徴とする。

【 0 0 1 3 】

かかる構成により、バーズイン制御部において検出される種々の特徴量に応じて、音声認識部におけるバーズイン機能の有効／無効を判断することが可能となり、雑音やエコー等に起因する不測のバーズインを抑止することができるとともに、雑音区間をユーザ音声区間と判断することに起因する誤認識も回避することができ、結果的に音声対話システムの誤動作を未然に防止することが可能となる。

【 0 0 1 4 】

また、本発明にかかる音声対話システムは、1つ以上の特徴量のうち少なくとも1つが雑音特徴量であり、雑音特徴量が所定のしきい値を越えた場合にバージイン機能を無効にすることが好ましい。雑音信号のパワー等の大小に応じてバージインを行うべきか否かを判断できるからである。

【 0 0 1 5 】

また、本発明にかかる音声対話システムは、1つ以上の特徴量のうち少なくとも1つがユーザ音声のS/Nであり、S/Nが所定のしきい値を越えた場合にバージイン機能を有効にすることが好ましい。雑音信号のS/Nの大小に応じてバージインを行うべきか否かを判断できるからである。

【 0 0 1 6 】

また、本発明にかかる音声対話システムは、1つ以上の特徴量のうち少なくとも1つが外部入力からの入力信号から検出されたユーザの位置情報であり、位置情報と環境情報に基づいてユーザがいる地点の環境評価値を算出し、環境評価値が所定のしきい値を越えた場合にバージイン機能を無効にすることが好ましい。ユーザの現在位置がどこであるかに基づいて、雑音信号が多い環境であるか否かを判断することができ、バージインを行うべきか否かを判断できるからである。

【 0 0 1 7 】

また、本発明にかかる音声対話システムは、音声対話部が、音声あるいは他の情報伝達手段の少なくとも1つ以上を用いて、バージイン機能の有効または無効の状態をユーザに通知することが好ましい。ユーザが音声出力の途中で入力することができるか否かを判断できるからである。

【 0 0 1 8 】

また、本発明は、上記のような音声対話システムの機能をコンピュータの処理ステップとして実行するソフトウェアを特徴とするものであり、具体的には、入力された音信号に対して音響的な信号処理を行う第一の工程と、音響的な信号処理後の音信号に含まれる音声の内容を認識する第二の工程と、音声の内容に基づいて、音声出力あるいは音声出力と他の情報伝達手段との併用によって、ユーザへ情報を伝達する第三の工程と、情報を伝達する途中に、第一の工程における入

力又は出力あるいは外部入力からの入力信号により、情報の伝達を中止するバージイン機能を有する第四の工程を含み、第四の工程において、第一の工程における入力又は出力あるいは外部入力からの入力信号から1つ以上の特徴量を検出し、1つ以上の特徴量に基づいてバージイン機能の有効又は無効を決定する音声対話方法並びにそのような工程を具現化するコンピュータ実行可能なプログラムであることを特徴とする。

【 0 0 1 9 】

かかる構成により、コンピュータ上へ当該プログラムをロードさせ実行することで、検出される種々の特徴量に応じて、バージイン機能を有効にするか否かを判断することが可能となり、雑音やエコー等に起因する不測のバージインを抑止することができるとともに、雑音区間をユーザ音声区間と判断することに起因する誤認識も回避することができ、結果的に音声対話システムの誤動作を未然に防止することができる音声対話システムを実現することが可能となる。

【 0 0 2 0 】

【発明の実施の形態】

以下、本発明の実施の形態にかかる音声対話システムについて、図面を参照しながら説明する。図1は本発明の実施の形態にかかる音声対話システムの基本構成図である。

【 0 0 2 1 】

図1において、11は入力された音信号に対して音響的な信号処理を行う音響処理部を、12は入力された音信号に含まれる音声の内容を認識するとともに、合成音声等による音声出力を含む情報伝達の途中において、ユーザにより入力された音信号により、情報の伝達を中止するバージイン機能を提供する音声認識部を、13は音声出力あるいは音声出力と他の情報伝達手段との併用によって、ユーザへ情報を伝達する音声対話部を、それぞれ示している。

【 0 0 2 2 】

また、14はバージイン制御部を示しており、入力された音信号から1つ以上の特徴量を検出し、1つ以上の特徴量に基づいてバージイン機能の有効又は無効を決定するものである。

【 0 0 2 3 】

まず音響処理部 1 1 は、入力された音信号に対して音響的な信号処理を行うことにより、当該音信号に含まれている様々な雑音を抑圧し、音声認識部 1 2 における音声認識の精度を高めることになる。ここで雑音とは、例えば電話回線を用いたボイスポータルシステムの場合においては、環境雑音や電話回線網雑音等の雑音のほか、回線エコーや音響エコー等の音声の回り込み等に起因するエコー雑音が考えられる。

【 0 0 2 4 】

環境雑音とは、ユーザが電話をかけている周辺を取り巻く環境において発生する種々の音信号が重畳した雑音を意味する。また電話回線雑音とは、公衆回線網上、特に伝送路上や交換機上で発生する各種のノイズ信号が重畳した雑音を意味している。

【 0 0 2 5 】

さらに、回線エコーとは、ボイスポータルがユーザへ提供する音声出力が、電話回線上において反射することにより重畳する雑音を意味する。また音響エコーとは、例えばハンズフリー電話を使用した場合、ボイスポータルからの音声出力は電話機の送話器で拡声して再生されるため、再生された音声出力が受話器側へ筐体振動によって回り込んだり、あるいは居室内における天井や壁、床などで反射して受話器へ回り込むことで重畳する雑音を意味している。

【 0 0 2 6 】

音響処理部 1 1 において、入力された音信号に含まれている上述したような雑音を抑圧するには、広く知られているノイズキャンセラ技術を利用することが多い。ノイズキャンセラ技術の代表例としては、スペクトルサブトラクション法が挙げられる。スペクトルサブトラクション法では、雑音区間において定常性雑音の振幅スペクトルを推定し、雑音区間ならびに雑音が重畳した音区間から、推定した振幅スペクトルを除去することによって雑音の抑圧を行うものである。ただし、ノイズキャンセラ技術として、これに限定されるものではない。

【 0 0 2 7 】

また、音響処理部 1 1 において、入力された音信号に含まれている回線エコー

や音響エコーの抑圧には、広く知られているエコーキャンセラ技術を利用することが多い。エコーキャンセラ技術の代表例としては、適応フィルタが挙げられる。適応フィルタは、FIRフィルタと呼ばれるデジタルフィルタで構成されており、エコーの発生源となる音信号である参照音（本実施の形態では音声出力信号）と入力された音信号との二乗誤差を最小とするように適応フィルタ係数を随時自動更新しながら、エコー抑圧を行うフィルタである。

【0028】

次に、音声認識部12は、雑音を抑圧した後の音響処理部11の出力を入力としており、音響処理部11の出力の特徴量を所定のしきい値と比較することによって、音響処理部11の出力に含まれるユーザ音声信号区間を検出するように構成されている。

【0029】

例えば、考えられる検出方法として、所定のしきい値に対して、入力された音信号の振幅あるいはパワーがしきい値を超えている区間をユーザによる音声区間と、超えていない区間を雑音区間と判定するように構成することになる。なお、音響処理部11からの出力の特徴量については、後述するバージン制御部14において検出する特徴量と同一の特徴量であることが好ましいが、異なる特徴量であっても良い。

【0030】

また、音声認識部12は、種々の音声認識技術を用いることにより、検出したユーザによる音声区間のユーザ発話内容を認識する。そして、認識結果については、音声対話部13へ出力することになる。

【0031】

次に、音声対話部13は、音声認識部12における音声認識結果に基づいて、ユーザに対して所望の情報を出力する。出力方法（情報伝達方法）としては、合成音声等による音声出力あるいは音声出力とその他の情報伝達手段の組み合わせが考えられる。なお、「その他の情報伝達手段」とは、例えば画像、音、光、振動等による情報伝達手段を意味している。

【0032】

また、音声対話部 1 3 は、バージイン制御部 1 4 においてバージイン機能が有効であると判断されており、バージインの発生が通知された場合、ユーザへの情報伝達を即座に中断する。さらに、必要に応じて音声対話部 1 3 が、ユーザへの情報伝達区間（音声出力時間）に関する情報を音響処理部 1 1 に出力する。

【 0 0 3 3 】

次に、バージイン制御部 1 4 は、入力又は音響処理部 1 2 の出力、あるいは外部入力のうち、少なくとも 1 つの特徴量を検出することによって、音声認識部 1 2 においてバージインを行うか否か、すなわちバージイン機能を有効にするか、無効にするかを判断するものである。ここで特徴量とは、例えば雑音信号の特徴量、S/N、ユーザの位置情報等が考えられる。

【 0 0 3 4 】

また、バージイン制御部 1 4 はバージイン機能を備えている。バージイン機能が有効であると判断されている場合、音響処理部 1 1 の出力の特徴量と所定のしきい値とを比較することによって、当該バージイン機能がユーザにより入力された音信号の立ち上がり（音声入力の開始）を検出することになる。また、バージイン機能が有効であると判断されている場合には、音声対話部 1 3 に対してバージインの発生を通知する。

【 0 0 3 5 】

図 2 は、特徴量を雑音信号の特徴量とした場合のバージイン制御部 1 4 の構成図である。図 2 においては、ユーザにより入力された音信号、あるいは音響処理部 1 1 において処理された後の音声信号の、少なくとも 1 つを入力としている。

【 0 0 3 6 】

そして、かかる入力から雑音特徴量を検出する雑音特徴量抽出部 1 4 1 と、雑音特徴量抽出部 1 4 1 で抽出された特徴量を所定のしきい値と比較することにより、バージイン機能を有効にするか否かを判定し、判定結果を音声認識部 1 2 及び音声対話部 1 3、あるいは必要に応じて音響処理部 1 1 に出力するバージイン判定部 1 4 2 を備えている。

【 0 0 3 7 】

雑音特徴量抽出部 1 4 1 において検出される雑音特徴量としては、例えば雑音

の特徴量を数値化した雑音の振幅値、あるいはその平均値や変化量を用いることが考えられる。ただし、これらに限定されるものではなく、パワー値や振幅スペクトル値、あるいはパワースペクトル値を用いても良い。

【 0 0 3 8 】

また、バーズイン判定部 1 4 2 において、雑音特徴量と比較する所定のしきい値は、バーズイン機能がユーザにより入力された音信号の立ち上がり（音声入力の開始）を検出するために用いる音響処理部 1 1 の出力の特徴量と比較する所定のしきい値よりも小さな値を設定することが好ましい。そうしなければ、音声出力途上に何らかの音信号（雑音だけの場合であっても）が入力されるごとにバーズイン機能が有効になってしまうことから、使い勝手が悪くなってしまうからである。

【 0 0 3 9 】

また、雑音特徴量が当該所定のしきい値を超える場合には、雑音レベルが高く、不測のバーズインが発生する可能性が高いと判断され、バーズイン判定部 1 4 2 がバーズイン機能が無効にするよう判断することになる。なお、バーズイン判定部 1 4 2 における所定のしきい値は、固定値であっても良いし、条件に応じて動的に変化する可変値であっても良い。

【 0 0 4 0 】

このような構成にすることで、バーズイン制御部 1 4 において検出された雑音特徴量に応じて、バーズイン機能の有効／無効を判断することが可能となり、雑音に起因する不測のバーズインを抑止することができるとともに、雑音区間をユーザ音声区間と判断することに起因する誤認識も回避することができ、結果的に音声対話システムの誤動作を未然に防止することが可能となる。

【 0 0 4 1 】

同様に図 3 は、特徴量を S/N（Sound/Noise）とした場合のバーズイン制御部 1 4 の構成図である。図 3 に示すように、ユーザにより入力された音入力信号、あるいは音響処理部 1 1 において処理された後の音声信号の、少なくとも 1 つを入力としている。

【 0 0 4 2 】

そして、かかる入力から S/N を検出する S/N 抽出部 1 4 3 と、 S/N 抽出部 1 4 3 で抽出された S/N を所定のしきい値と比較することにより、バーズイン機能を有効にするか否かを判定し、判定結果を音声認識部 1 2 及び音声対話部 1 3、あるいは必要に応じて音響処理部 1 1 に出力するバーズイン判定部 1 4 2 を備えている。

【 0 0 4 3 】

また、バーズイン判定部 1 4 2 において、 S/N と比較する所定のしきい値は、バーズイン機能がユーザにより入力された音信号の立ち上がり（音声入力の開始）を検出するために用いる音響処理部 1 1 の出力の特徴量と比較する所定のしきい値よりも大きな値を設定することが好ましい。そうしなければ、音声出力途上にユーザ音声よりもレベルの高い非定常雑音が入力されるごとにバーズイン機能が有効になってしまうことから、使い勝手が悪くなってしまうからである。

【 0 0 4 4 】

また、 S/N が当該所定のしきい値を超えない場合には、ユーザにより入力される音信号の S/N が低く、レベルの高い非定常雑音による不測のバーズインが発生する可能性が高いと判断され、バーズイン判定部 1 4 2 がバーズイン機能を無効にするように判断することになる。なお、バーズイン判定部 1 4 2 における所定のしきい値は、固定値であっても良いし、条件に応じて動的に変化する可変値であっても良い。

【 0 0 4 5 】

このような構成にすることで、バーズイン制御部 1 4 において検出された S/N に応じて、バーズイン機能の有効／無効を判断することが可能となり、ユーザ音声よりもレベルの高い非定常雑音に起因する不測のバーズインを抑止することができるとともに、雑音区間をユーザ音声区間と判断することに起因する誤認識も回避することができ、結果的に音声対話システムの誤動作を未然に防止することが可能となる。

【 0 0 4 6 】

同様に図 4 は、特徴量をユーザの位置情報とした場合のバーズイン制御部 1 4 の構成図である。図 4 に示すように、ユーザにより入力される音声とは相違する

、別の外部入力を入力としている。例えばGPSからの位置情報等である。

【0047】

そして、かかる入力からユーザの位置情報を検出する位置情報抽出部144と、位置情報抽出部144で抽出されたユーザの位置情報に基づいて周辺環境の評価値を算出する環境評価値算出部145と、環境評価値算出部145において算出された評価値を所定のしきい値と比較することにより、バージイン機能を有効にするか否かを判定し、判定結果を音声認識部12及び音声対話部13、あるいは必要に応じて音響処理部11に出力するバージイン判定部142を備えている。

【0048】

図4に示す位置情報抽出部144は、外部入力から、例えばGPS等によって得られるユーザの位置座標を取得することになる。そして、環境評価値算出部145は、位置情報抽出部144により抽出されたユーザの位置座標と、事前に登録しておいたマップ情報等の環境情報とを照合することによって所定の環境評価値を算出し、バージイン判定部142では、算出された環境評価値に基づいて、バージイン機能を有効にするか否かを判断することになる。ここでマップ情報とは、例えば環境雑音が多いと予測されるエリア（例えば幹線主要道路付近等）を高雑音エリア、環境雑音の少ないエリア（閑静な住宅街等）を低雑音エリアとして登録したエリア情報データベース等が考えられる。もちろん、これに限定されるものではない。

【0049】

また、環境評価値としては、周辺環境における環境雑音の期待値等が考えられる。すなわち、ユーザの位置座標から現在ユーザが位置する周辺の環境を推定することができ、例えば周辺環境における環境雑音の期待値が所定のしきい値を超える場合には、環境雑音の大きな場所にいるものと判断され、不測のバージインが発生する可能性が高いと判断でき、バージイン判定部14においてバージイン機能を無効にするように判断することになる。なお、バージイン判定部142における所定のしきい値は、固定値であっても良いし、条件に応じて動的に変化する可変値であっても良い。

【 0 0 5 0 】

このような構成にすることで、バージイン制御部 1 4 において検出されたユーザの位置情報に基づいて算出された環境評価値に応じて、バージイン機能の有効／無効を判断することが可能となり、ユーザの周辺環境による環境雑音に起因する不測のバージインを抑止することができるとともに、雑音区間をユーザ音声区間と判断することに起因する誤認識も回避することができ、結果的に音声対話システムの誤動作を未然に防止することが可能となる。

【 0 0 5 1 】

また、上述した 3 つの特徴量抽出手段を組み合わせても良い。すなわち、バージイン制御部 1 4 に、雑音特徴量抽出部 1 4 1 と、S/N 抽出部 1 4 3 と、位置情報抽出部 1 4 4 とを設け、それぞれ雑音特徴量としきい値 A、S/N としきい値 B、ユーザの位置情報に基づく評価値としきい値 C の比較を行い、バージイン機能を有効にするか否かを判断することになる。なお、本実施の形態においては、3 種類の特徴量抽出手段を設ける場合について説明しているが、3 種類以上設けても良く、また 3 種類のうちのいずれか 2 種を設ける構成であっても良い。

【 0 0 5 2 】

このように異なる種類の特徴量抽出手段を併用することで、バージイン機能を有効にするべきか否かの判断精度の向上が期待でき、不測のバージインの抑制、及び雑音区間をユーザ音声区間と判断することに起因する誤認識の抑制を、より効果的に行うことが期待できる。

【 0 0 5 3 】

以下、特徴量として雑音信号の特徴量を用いる場合を例に挙げて、具体的な処理について図面を参照しながら説明する。図 5 は、本発明の実施の形態にかかる音声対話システムであって、初期状態において音声認識部 1 2 におけるバージイン機能が無効とされている場合の入出力状態及びバージイン機能の有効／無効の状態を示す図である。

【 0 0 5 4 】

図 5 において、バージイン制御部 1 4 は、ユーザによる音声入力区間を除く区間において雑音特徴量を検出し、雑音特徴量に応じてバージイン機能が有効であ

るか否かを判断することになる。

【 0 0 5 5 】

ここで、バーズイン機能が初期状態において無効に設定されている場合、少なくともユーザが音声対話システムへアクセスを開始した直後から、音声対話部 1 3 における音声出力による情報伝達が終了するまでの区間、すなわち図 5 における区間①-A 及び①-B においては、雑音特徴量に関係なくバーズイン機能を無効に設定するようにする。すなわち、ユーザが音声対話システムへアクセスを開始した直後は音声対話システムにとって雑音レベルが未知であることから、雑音等による不測のバーズイン発生防止を目的として、区間①-A 及び①-B においてバーズイン機能を無効とするものである。

【 0 0 5 6 】

また、バーズイン制御部 1 4 で検出する雑音特徴量は、区間①における区間①-A 及び①-C で検出することになる。当該区間においては、環境雑音や回線網雑音等の雑音特徴量を確実に検出することができるからである。一方、区間①-B では、環境雑音や回線網雑音等の雑音に加えて、残留回線エコーや残留音響エコー等の残留エコーを重ね合わせた雑音特徴量を検出することになるため、正確さを欠くおそれも残されている。

【 0 0 5 7 】

なお、雑音特徴量を検出する区間としては区間①全体に限定されるものではなく、例えば区間①-A、①-B、及び①-C のいずれか 1 つの区間であっても良く、あるいは区間①-A、①-B、及び①-C の一部区間に限定しても良い。

【 0 0 5 8 】

そして、バーズイン制御部 1 4 は、区間①における雑音特徴量と所定のしきい値との比較を行い、雑音特徴量が所定のしきい値を超えた場合には、雑音レベルが高く、不測のバーズインが発生する可能性が高いと判断して、バーズイン機能を無効にするよう設定することになる。ただし、図 5 では、雑音特徴量が所定のしきい値を超えていない場合について示しており、バーズイン制御部 1 4 はバーズイン機能を有効にするよう設定している。

【 0 0 5 9 】

なお、図 5 ではバージン制御部 1 4 が区間①-B 終了後に音声認識部 1 2 におけるバージン機能を有効／無効のいずれかに設定するようにしているが、バージン機能の設定タイミングはこれに限定されるものではなく、区間①における任意の位置において有効／無効のいずれかに設定するようにすれば良い。

【 0 0 6 0 】

次に、バージン制御部 1 4 は、ユーザによる音声入力区間では特徴量検出を行わず、バージン機能は区間①-B 終了後に設定した状態、すなわち有効である状態を保持している。ユーザによる音声入力が行われている間は、雑音にユーザによる音声入力が重畳するため、バージン制御部 1 4 における雑音特徴量は、雑音の特徴量のみを検出することが困難であることから、雑音の特徴量のみを検出できる区間①-C においてバージン機能を有効にするためである。

【 0 0 6 1 】

そして、バージン機能が有効と設定されている場合、ユーザによる音声入力の終了後から、音声対話部 1 3 において音声出力が開始され、所定の時間を経過した区間②において雑音特徴量を検出し、ある所定のしきい値を超えた場合、図 5 に示すようにバージン機能を無効にするよう設定する。

【 0 0 6 2 】

なお、雑音特徴量を検出する区間を②としているが、例えば区間②-A あるいは②-B のいずれかに限定しても良く、また区間②-A 及び②-B の一部区間に限定しても良い。区間②において、音声対話部 1 3 から音声出力が開始され、所定の時間を経過した区間②-B は、例えば音声対話部 1 3 から音声出力が開始されてからユーザによる音声入力が始まる（バージンの場合も含む）直前までとしても良く、予め所定の時間長さを設定しておいても良い。ただし、後者の場合、音声対話部 1 3 から音声出力が開始され、ユーザによる音声入力によるバージンが発生する区間より短い区間を予測して設定することになる。

【 0 0 6 3 】

また、図 5 に示すように、状況に応じて動的にバージン機能の有効／無効を切り替えるのではなく、例えば図 6 に示すように初期状態においてはバージン機能が無効になるよう設定されており、条件を具備した場合に 1 回のみ有効にな

るように切り替える方法や、あるいは図 7 に示すように初期状態においてはバージイン機能が有効になるよう設定されており、条件を具備した場合に 1 回のみ無効になるように切り替える方法も考えられる。

【 0 0 6 4 】

ここで、初期状態においてバージイン機能が無効になるよう設定することは、ユーザが音声対話システムにアクセスした際、未知レベルの様々な雑音に対し、不測のバージイン発生を防止する効果があるからである。

【 0 0 6 5 】

そして、図 6 に示すように、バージイン制御部 1 4 は、区間①における雑音特徴量が、例えばある所定のしきい値を超えない場合には、雑音レベルが低く、不測のバージインが発生する可能性が低いと判断して、バージイン機能が有効となるよう設定し、有効状態のまま固定することになる。なお、図 6 では区間①-B 終了後にバージイン機能を有効にするよう設定しているが、特にこの位置に限定されるものではなく、区間①の任意の位置において、バージイン機能を有効にするよう設定しても良い。

【 0 0 6 6 】

また、初期状態においてバージインを有効に設定することは、音声対話システムからの音声出力による情報伝達の途中であっても、ユーザの所望のタイミングで割り込みをかけられるというメリットがある。

【 0 0 6 7 】

そして、バージイン制御部 1 4 は、区間①における雑音特徴量が、例えばある所定のしきい値を超える場合には、雑音レベルが高く、不測のバージインが発生する可能性が高いと判断して、バージイン機能を無効となるよう設定し、無効状態のまま固定することになる。図 7 では、区間①-B 終了後にバージイン機能を無効にするよう設定しているが、特にこの位置に限定されるものではなく、区間①の途中でバージイン機能が無効になるよう設定しても良い。

【 0 0 6 8 】

このように、初期状態から 1 回のみ状態を切り替えることによって、結果的に音声対話システムの誤動作を防止することが可能であるとともに、バージイン制

御部 1 4 の構成を単純化することができ、音声対話システム全体の低コスト化を図ることも可能となる。

【 0 0 6 9 】

なお、前述した適応フィルタを用いる場合には、以下のような問題に留意する必要がある。すなわち、一般に適応フィルタは、入力される音信号と参照音信号とが相関が低いことを前提としており、適応アルゴリズムを用いた適応処理によって、逐次 F I R フィルタ係数を更新するものである。しかし、実際の入力音はエコー音だけではなく、場合によってはユーザによる音声自体が重畳する場合もある。この場合、同一の音声の回り込み等であることから相関が高くなってしまうため、音響処理部 1 2 が適応フィルタを備えている場合には適応処理によって誤ったフィルタ係数に更新してしまい、エコー抑圧効果が低減するとともに、ユーザにより入力された音声自体を歪ませることになり、ユーザにより入力された音声の内容を認識できないという問題である。

【 0 0 7 0 】

したがって、バージン機能が無効となるように設定されている場合、適応フィルタによる適応処理区間を、音声対話部 1 3 から音声出力している音声情報伝達区間のみに限定し、非音声情報伝達区間では適応処理を行わないようにする必要がある。

【 0 0 7 1 】

このようにすることで、音声対話部 1 3 が音声出力している情報伝達区間においてはバージンが発生しないため、エコー音にユーザによる入力音声为重畳することがなく、適応処理及び適応フィルタ係数の更新が正常に行われるとともに、ユーザによる入力音声自体が歪むこともなく、雑音であるエコー音のみを抑圧することが可能となる。

【 0 0 7 2 】

さらに、音響処理部 1 1 が適応フィルタによって構成されている場合、バージン機能が有効となるよう設定されていると、音声対話部 1 3 からの音声出力開始から所定の時間までの区間に限定して、適応フィルタによる適応処理とそれにかかる適応フィルタ係数の更新を行うようにする。

【 0 0 7 3 】

ここで「所定の時間」とは、長くとも音声対話部 1 3 からの音声出力の開始時刻から、ユーザによる音声入力によるバージイン開始時刻までの時間であれば良く、あるいは事前に設定した時間であっても良い。ただし、後者の場合、ユーザによる音声入力によるバージイン開始より短い時間を推定して設定する必要がある。

【 0 0 7 4 】

このようにすることで、音響処理部 1 1 における適応処理と適応フィルタ係数の更新が、音声対話部 1 3 が音声出力を開始してから長くともユーザによるバージインが発生するまでの区間に限定して行われることから、雑音であるエコー音にユーザ音声为重畳することが無く、適応処理と適応フィルタ係数の更新が正常に行われるとともに、ユーザによる入力音声自体が歪むこともなく、雑音であるエコー音のみを抑圧することが可能となる。

【 0 0 7 5 】

次に、本発明の実施の形態にかかる音声対話システムを実現するプログラムの処理の流れについて説明する。図 8 に本発明の実施の形態にかかる音声対話システムを実現するプログラムの処理の流れ図を示す。

【 0 0 7 6 】

図 8 において、まずユーザにより入力された音声信号を受信し（ステップ S 8 0 1）、音響的な信号処理を行うことにより、当該音声信号に含まれている様々な雑音を抑圧する（ステップ S 8 0 2）。

【 0 0 7 7 】

次に、雑音抑圧後の音声信号について、その特徴量を抽出し（ステップ S 8 0 3）、抽出された特徴量を所定のしきい値と比較する（ステップ S 8 0 4）。特徴量が所定のしきい値よりも大きい場合には（ステップ S 8 0 4 : Y e s）、当該区間をユーザによる音声信号区間として検出することになる（ステップ S 8 0 5）。特徴量が所定のしきい値以下である場合には（ステップ S 8 0 4 : N o）、当該区間をユーザによる雑音信号区間として検出することになる（ステップ S 8 0 8）。

【 0 0 7 8 】

ユーザによる音声信号区間として検出された区間については、種々の音声認識技術を用いることによって、発話内容を認識する（ステップ S 8 0 6）。そして、認識結果に対応してユーザに伝達すべき情報が、ユーザに対して音声出力によって伝達される（ステップ S 8 0 7）。

【 0 0 7 9 】

一方、雑音信号区間として検出された区間については、少なくとも 1 種類の特徴量を抽出する（ステップ S 8 0 9）。そして、抽出された特徴量を所定のしきい値と比較して（ステップ S 8 1 0）、所定のしきい値より大きい場合には（ステップ S 8 1 0 : Y e s）バーズイン機能を有効にし（ステップ S 8 1 1）、所定のしきい値以下である場合には（ステップ S 8 1 0 : N o）バーズイン機能を無効にする（ステップ S 8 1 5）。なお、特徴量の種類によっては、所定のしきい値を超えていない場合にバーズイン機能を有効にし、超えている場合に無効にするよう制御される。

【 0 0 8 0 】

次に、バーズイン機能が有効である場合には（ステップ S 8 1 1）、継続して入力されている雑音抑圧後の音声信号について、その特徴量を所定のしきい値と比較する等して、ユーザによる音声入力の開始を検出する（ステップ S 8 1 2）。同時に、音声対話部 1 3 に対してバーズインの発生を通知する（ステップ S 8 1 3）。

【 0 0 8 1 】

そして、バーズイン機能が有効であると判断されており（ステップ S 8 1 1）、かつバーズインの発生が通知された場合（ステップ S 8 1 3）、ユーザへの音声出力が中断される（ステップ S 8 1 4）。

【 0 0 8 2 】

なお、本発明の実施の形態にかかる音声対話システムを実現するプログラムは、図 9 に示すように、C D - R O M 9 2 - 1 やフレキシブルディスク 9 2 - 2 等の可搬型記録媒体 9 2 だけでなく、通信回線の先に備えられた他の記憶装置 9 1 や、コンピュータ 9 3 のハードディスクや R A M 等の記録媒体 9 4 のいずれに記

憶されるものであっても良く、プログラム実行時には、プログラムはローディングされ、主メモリ上で実行される。

【 0 0 8 3 】

また、本発明の実施の形態にかかる音声対話システムにより用いられる各特徴量に対応したバージン機能の有効／無効判断のためのしきい値データ等についても、図 9 に示すように、CD-ROM 9 2 - 1 やフレキシブルディスク 9 2 - 2 等の可搬型記録媒体 9 2 だけでなく、通信回線の先に備えられた他の記憶装置 9 1 や、コンピュータ 9 3 のハードディスクや RAM 等の記録媒体 9 4 のいずれに記憶されるものであっても良く、例えば本発明にかかる音声対話システムを利用する際にコンピュータ 9 3 により読み取られる。

【 0 0 8 4 】

【発明の効果】

以上のように本発明にかかる音声対話システムによれば、バージン制御部において検出される種々の特徴量に応じて、バージン機能の有効／無効を判断することが可能となり、雑音やエコー等に起因する不測のバージンを抑止することができるとともに、雑音区間をユーザ音声区間と判断することに起因する誤認識も回避することができ、結果的に音声対話システムの誤動作を未然に防止することが可能となる。

【図面の簡単な説明】

【図 1】 本発明の実施の形態にかかる音声対話システムの構成図

【図 2】 本発明の実施の形態にかかる音声対話システムにおけるバージン制御部の構成図

【図 3】 本発明の実施の形態にかかる音声対話システムにおけるバージン制御部の構成図

【図 4】 本発明の実施の形態にかかる音声対話システムにおけるバージン制御部の構成図

【図 5】 本発明の実施の形態にかかる音声対話システムにおけるバージン機能の有効／無効制御の説明図

【図 6】 本発明の実施の形態にかかる音声対話システムにおけるバージン

機能の有効／無効制御の説明図

【図 7】 本発明の実施の形態にかかる音声対話システムにおけるバージイン

機能の有効／無効制御の説明図

【図 8】 本発明の実施の形態にかかる音声対話システムにおける処理の流れ
図

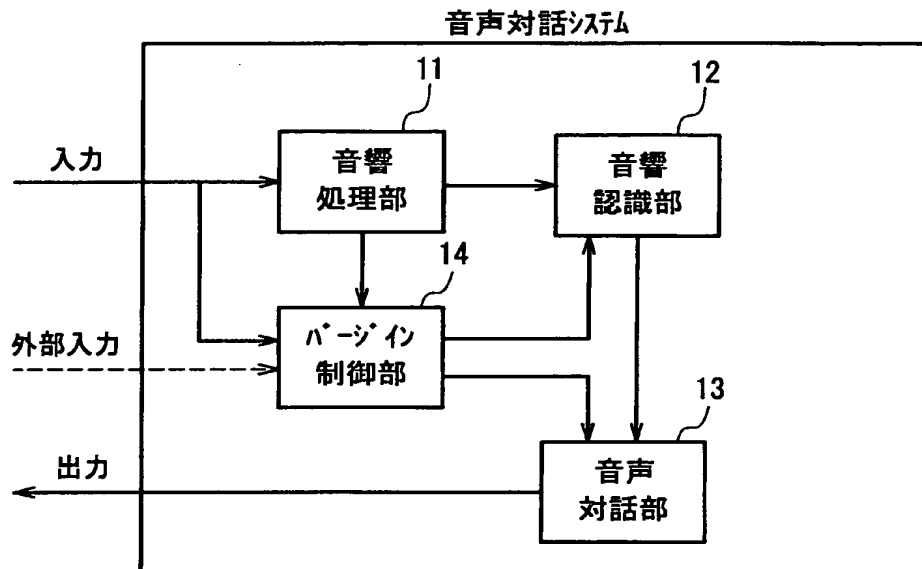
【図 9】 コンピュータ環境の例示図

【符号の説明】

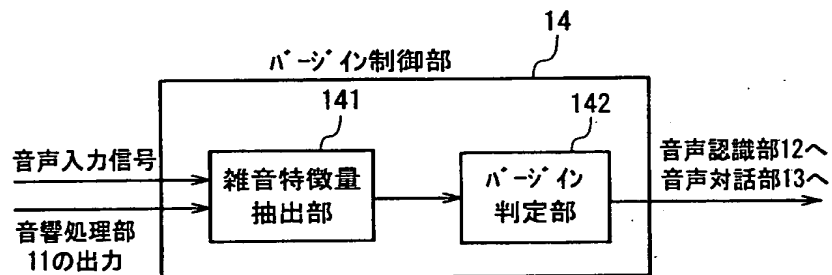
- 1 1 音響処理部
- 1 2 音声認識部
- 1 3 音声対話部
- 1 4 バージイン制御部
- 9 1 回線先の記憶装置
- 9 2 C D - R O M や フ レ キ シ ブ ル デ ィ ス ク 等 の 可 搬 型 記 録 媒 体
- 9 2 - 1 C D - R O M
- 9 2 - 2 フ レ キ シ ブ ル デ ィ ス ク
- 9 3 コ ン ピ ュ ー タ
- 9 4 コ ン ピ ュ ー タ 上 の R A M / ハ ー ド デ ィ ス ク 等 の 記 録 媒 体
- 1 4 1 雑音特徴量抽出部
- 1 4 2 バージイン判定部
- 1 4 3 S / N 抽出部
- 1 4 4 位置情報抽出部
- 1 4 5 環境評価値算出部

【書類名】 図面

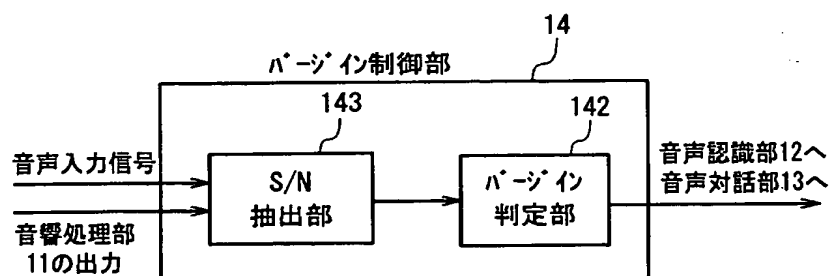
【図 1】



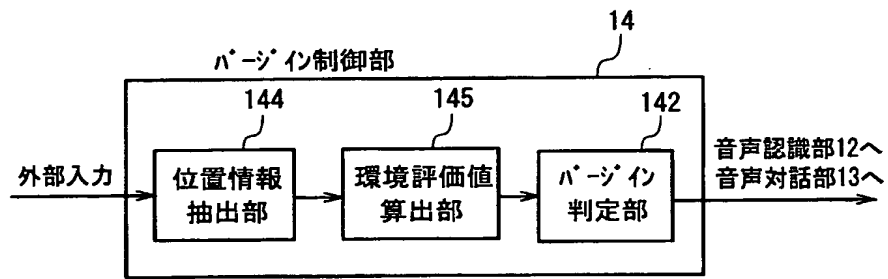
【図 2】



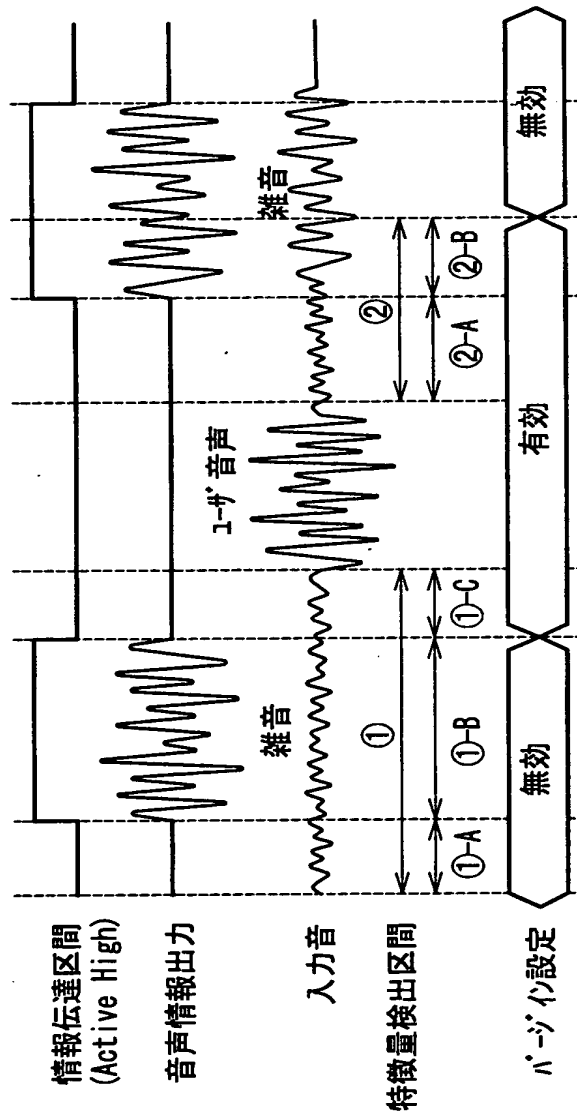
【図 3】



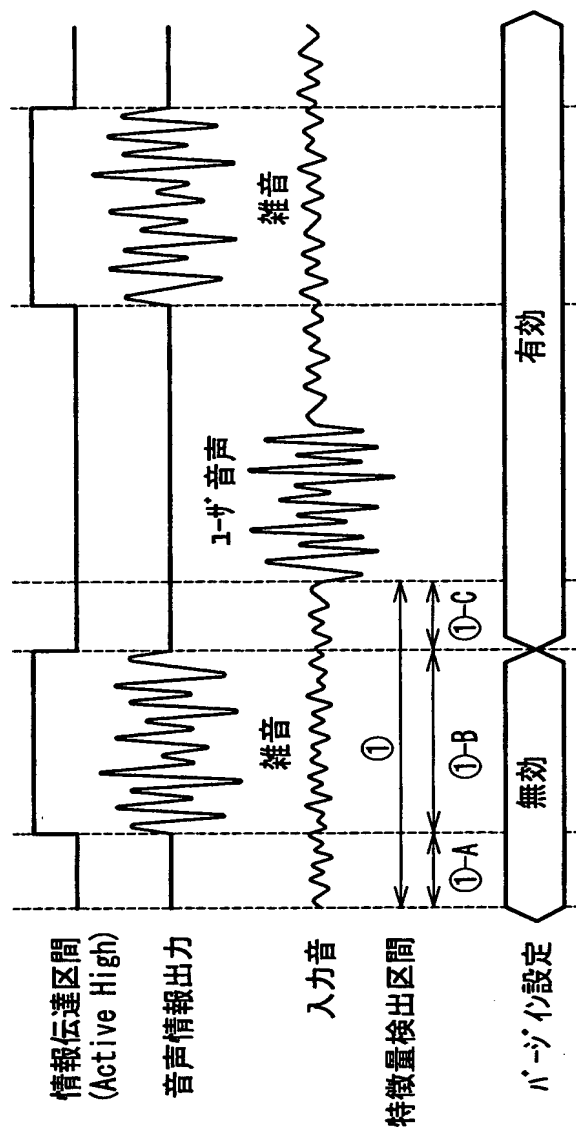
【図 4】



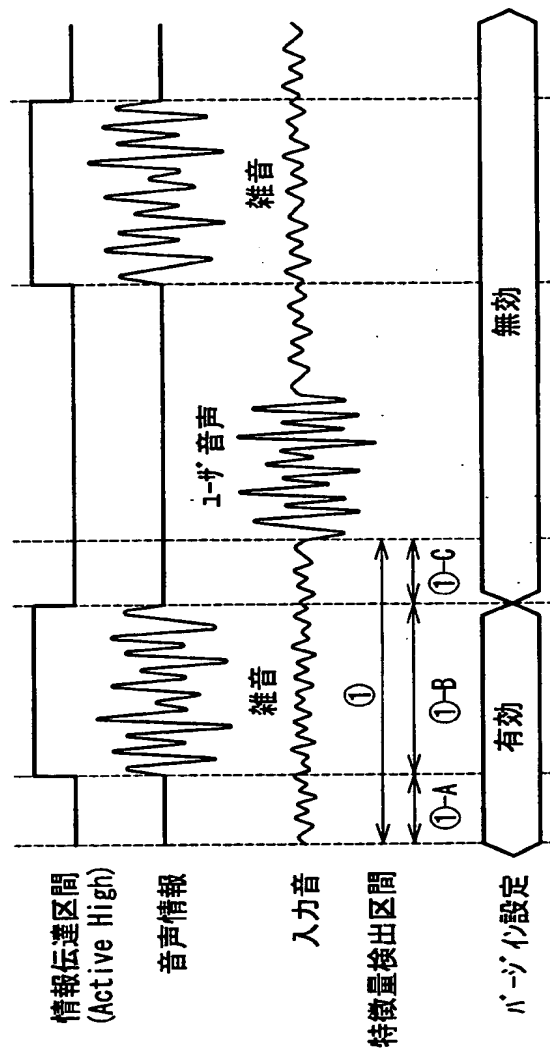
【図 5】



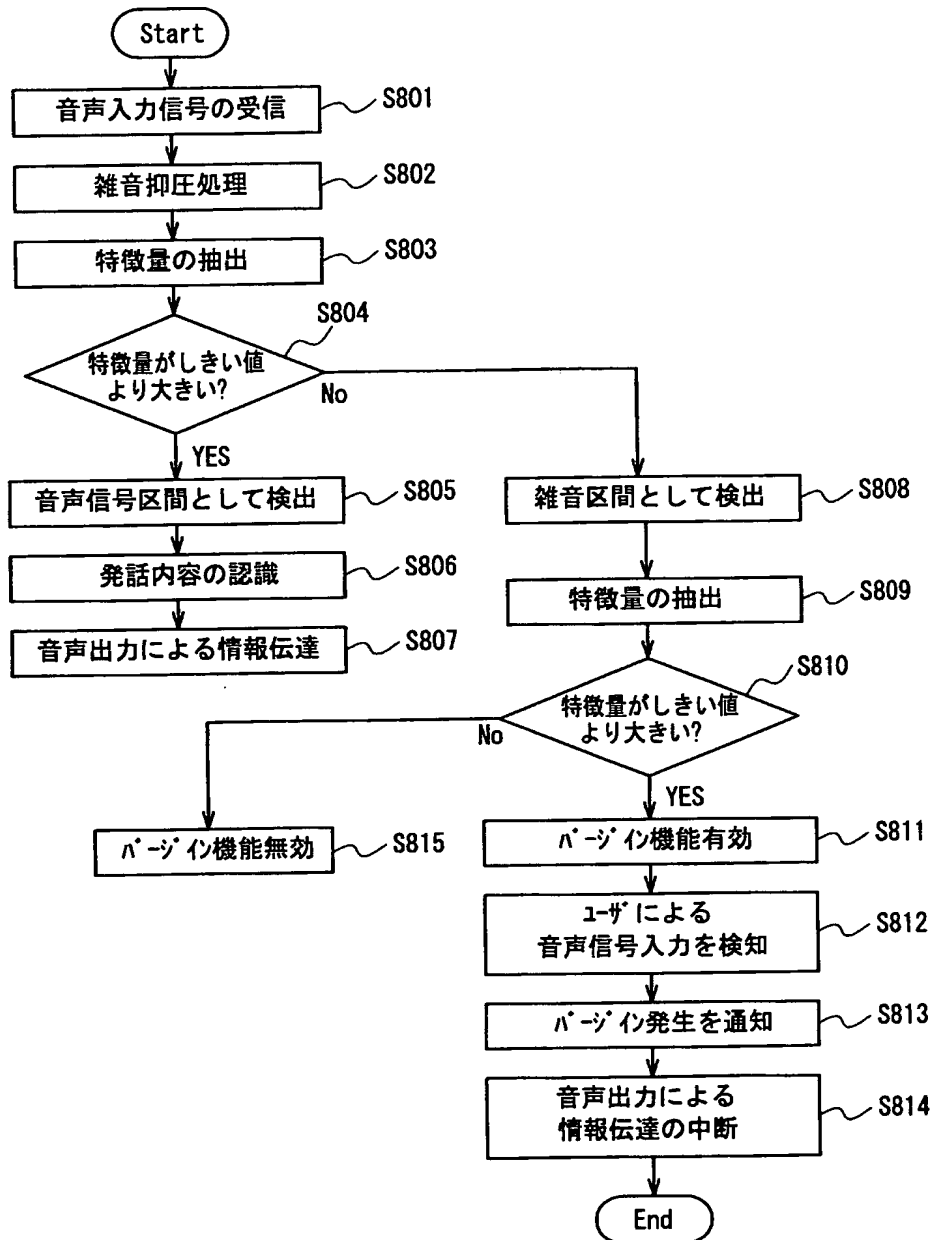
【図6】



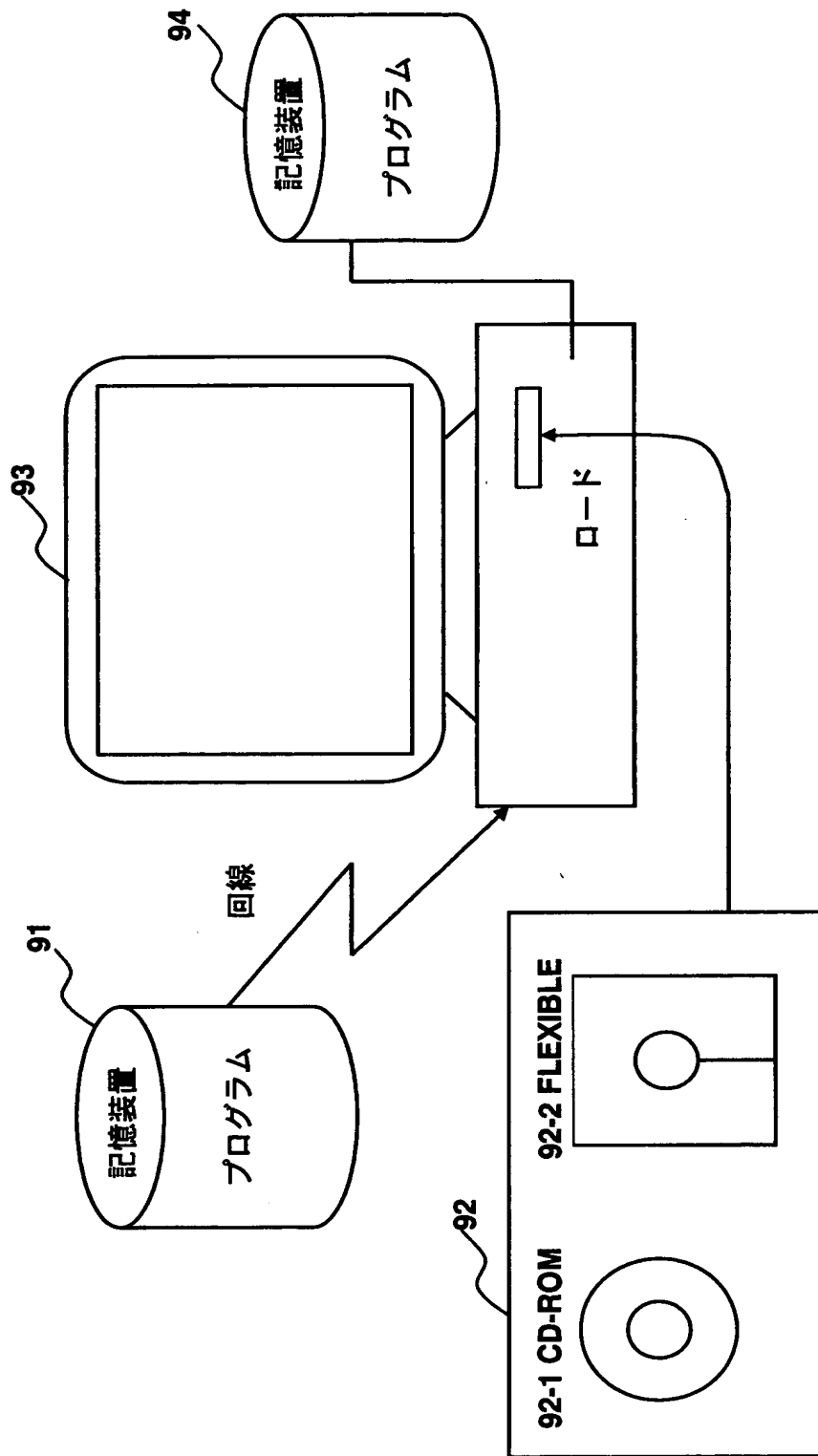
【図 7】



【図 8】



【図9】



【書類名】 要約書

【要約】

【課題】 様々なエコーや雑音に起因する不測のバージインと、それに伴う誤認識によるシステムの誤動作の両者を抑止することで、ユーザにとって利便性の高い音声対話システム及び方法を提供する。

【解決手段】 入力された音信号に対して音響的な信号処理を行い、音響的な信号処理後の音信号に含まれる音声の内容を認識し、音声の内容に基づいて、音声出力あるいは音声出力と他の情報伝達手段との併用によって、ユーザへ情報を伝達するとともに、情報を伝達する途中に、音響的信号処理の入力又は出力あるいは外部入力からの入力信号により、情報の伝達を中止するバージイン機能を有する音声対話システムであって、音響的信号処理の入力又は出力あるいは外部入力からの入力信号から1つ以上の特徴量を検出し、1つ以上の特徴量に基づいてバージイン機能の有効又は無効を決定する。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [000005223]

1. 変更年月日	1996年 3月26日
[変更理由]	住所変更
住 所	神奈川県川崎市中原区上小田中4丁目1番1号
氏 名	富士通株式会社